Published online 29 November 2012 *Nucleic Acids Research, 2013, Vol. 41, Database issue* ***D285–D294***
doi:10.1093/nar/gks1238

YM500: a small RNA sequencing (smRNA-seq) database for microRNA research

Wei-Chung Cheng[1,2], I-Fang Chung[3], Tse-Shun Huang[4], Shih-Ting Chang[3], Hsing-Jen Sun[3], Cheng-Fong Tsai[3], Muh-Lii Liang[1], Tai-Tong Wong[1,2,5,*] and Hsei-Wei Wang[2,3,4,6,7,*]

[1]Division of Pediatric Neurosurgery, Neurological Institute, Taipei Veterans General Hospital, Taipei 11217, [2]VGH-YM Genomic Research Center, [3]Institute of Biomedical Informatics, [4]Institute of Microbiology and Immunology, National Yang-Ming University, Taipei 11221, [5]Division of Neurosurgery, Department of Surgery, Cheng Hsin Hospital, Taipei 11220, Taiwan, [6]Cancer Research Center, National Yang-Ming University, Taipei 11221, Taipei and [7]Department of Education and Research, Taipei City Hospital, Taipei 10341, Taiwan

Received August 14, 2012; Revised and Accepted November 1, 2012

ABSTRACT

MicroRNAs (miRNAs) are small RNAs ~22 nt in length that are involved in the regulation of a variety of physiological and pathological processes. Advances in high-throughput small RNA sequencing (smRNA-seq), one of the next-generation sequencing applications, have reshaped the miRNA research landscape. In this study, we established an integrative database, the YM500 (http://ngs.ym.edu.tw/ym500/), containing analysis pipelines and analysis results for 609 human and mice smRNA-seq results, including public data from the Gene Expression Omnibus (GEO) and some private sources. YM500 collects analysis results for miRNA quantification, for isomiR identification (incl. RNA editing), for arm switching discovery, and, more importantly, for novel miRNA predictions. Wetlab validation on >100 miRNAs confirmed high correlation between miRNA profiling and RT-qPCR results ($R = 0.84$). This database allows researchers to search these four different types of analysis results via our interactive web interface. YM500 allows researchers to define the criteria of isomiRs, and also integrates the information of dbSNP to help researchers distinguish isomiRs from SNPs. A user-friendly interface is provided to integrate miRNA-related information and existing evidence from hundreds of sequencing datasets. The identified novel miRNAs and isomiRs hold the potential for both basic research and biotech applications.

INTRODUCTION

MicroRNAs (miRNAs) are a family of small RNAs, ~22 nt in length, that act as key post-transcriptional regulators of gene expression, modulating the translational efficiency and/or the stability of target mRNAs. Small RNA sequencing (smRNA-seq), one of the next-generation sequencing (NGS) applications, enables detection and profiling of miRNAs with particularly high levels of sensitivity and accuracy (1). Furthermore, smRNA-seq allows discovery of previously uncharacterized miRNA species and has revealed unexpected complexity among miRNAs. These smRNA-seq discoveries include not only novel miRNAs but also a series of miRNA variants, termed isomiRs (2). In recent years, smRNA-seq datasets have grown rapidly and have been deposited in public databases such as the Gene Expression Omnibus (GEO) (3) and ArrayExpress (4). Exploration of these massive datasets, however, remains a daunting challenge, and an integrative meta-analysis of all smRNA-seq datasets has not yet been well performed.

For the detection and identification of novel miRNAs, smRNA-seq is very promising because it is not as time-consuming and expensive as small RNA cloning methods (5). Many sequencing software tools have been developed to identify novel miRNAs. Li *et al*. evaluated eight software tools, namely miRDeep (6), miRanalyzer (7), miRTRAP (8), MIReNa (9), mirTools (10), DSAP (11), miRNAkey (12) and mireap (13), based on their common features and key algorithms, and recommended the best tools in predicting novel miRNAs for different data types (5).

IsomiRs are commonly reported in deep sequencing studies (14–27) and are unlikely to be due simply to degradation or sequencing errors (25,28). These variants have

*To whom correspondence should be addressed. Tel: +886 2 2826 7109; Fax: +886 2 2821 2880; Email: hwwang@ym.edu.tw
Correspondence may also be addressed to Tai-Tong Wong. Tel: +886 2 2826 4545; Fax: +886 2 2826 4533; Email: ttwong@vghtpe.gov.tw

been reported to be biologically relevant and functionally cooperative partners of canonical miRNAs (14,21). The variations present in IsomiRs can be grouped into three types: editing (nucleotide substitution), trimming and addition (29). The latter two types cause 5′ and 3′ end-length heterogeneity of miRNAs. Editing is a consequence of adenosine or cytidine deaminase activities and causes nucleotide changes at different positions of the mature miRNAs (15,16,28,30–33). It has previously been shown that several miRNAs, edited in the seed sequence and with an increased level of editing throughout development, result in diversifying target recognition (16). Trimming results in the shorter mature miRNAs compared with the canonical ones. The 3′-to-5′ exoribonuclease Nibbler has also been reported to control 3′ end processing of miRNAs in Drosophila (34,35). Non-template nucleotide additions at the 3′ end of miRNAs have been reported as the common form of miRNA enzymatic modification (18,21,28) and can influence miRNA stability (36) and the efficiency of target repression (37). It has further been revealed that the frequency of 3′ addition to specific miRNAs changes with differentiation of human embryonic stem cells (18). Several enzymes, such as MTPAP, PAPD4, PAPD5, ZCCHC6, ZCCHC11 and TUT1, have been reported to govern 3′ nucleotide addition to miRNAs (18,36–38).

The miR/miR* nomenclature has been used to represent the dominant and minor mature products of precursor miRNAs. However, several studies have reported that the arm that makes the dominant product can change in different tissues, stages and species (14,39–43). Such changes have been called 'arm switching' and are likely to be general (39). For example, Grimson *et al.* have reported an instance of developmental arm switching between the embryonic and adult stages of sponges for miR-2015 (43). Cloonan *et al.* also listed several miRNAs whose dominant arm changes over different human tissues (14). Therefore, in release 19 of the miRBase database (miRBase R19), the nomenclature for mature miRNAs now designates them as -5p and -3p, rather than miR/miR*, in all species.

In this study, we present the YM500 database, which includes pipelines for miRNA quantification, isomiR identification, arm switching discovery and novel miRNA prediction from smRNA-Seq data. YM500 contains the results of meta-analysis from hundreds of public smRNA-Seq datasets and dozens of in-house ones. YM500 aims to provide researchers with integrated miRNA-related information with various graphical visualization pages from hundreds of sequencing datasets via a user-friendly web interface.

MATERIALS AND METHODS

Data collection and pre-processing

As shown under *Pre-processing* in Figure 1, there are 609 Illumina smRNA-Seq datasets, including 468 human and 141 mouse ones, in YM500. 34 out of 609 are in-house datasets, and the others are from the GEO public repository. All in-house data were deposited in the GEO database with an accession number of GSE39841. Detailed information for the datasets is provided in Supplementary Table S1. 3′ adaptors of the FASTQ raw data were trimmed, and the trimmed data were collapsed by the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit). A QC report, including length distribution, box plot of phred quality scores, etc., was then generated. For the public datasets containing only sequences and read counts, we transformed them into FASTA format and examined their length distribution. It has been suggested that if the smRNA-seq reads are abundant in 19∼24 nt, they are good for miRNA analysis (44). Thus, datasets which did not fit this criterion were discarded. For all datasets, reads shorter than 17 nt or longer than 30 nt were filtered out.

Prediction of novel miRNAs and their downstream targets

It has been noted that there is no evidence that the miRNAs may represent fragments of mRNAs or other known RNA types (45). As shown in the *Novel miRNA Prediction* module in Figure 1, before predicting novel miRNAs, we used Bowtie (46) with options: -v 0 -f –norc to filter out reads that map to known miRNA precursors (miRBase) (45), other functional RNAs (Rfam) (47) and mRNAs (RefSeq) (48). Then, we adopted three prediction tools, namely miRDeep2, miReap and miRanalyzer, for predicting novel miRNAs. All of the prediction results were merged and unified to remove redundant records. According to our experience, there are still some predicted novel miRNAs that were mapped to known transcripts. We further applied BLAST to remove reads which were fully mapped to RefSeq or Rfam with identity >90%. To get a more reliable result, we also used miReNa as a second filter to filter out those that do not satisfy numerical Criteria I–V to describe a pre-miRNA in miReNa. About two-thirds of the putative miRNAs could not fulfil miReNa criteria and were filtered out. Finally, we filtered out the putative miRNAs located in exon regions (defined by RefSeq). These filtrations aimed to reduce the false positive rate. Filtration results are shown in Supplementary Figure S1. Most of the putative novel miRNAs that were predicted by multiple algorithms were preserved. There are 90 and 637 putative novel miRNAs predicted by three algorithms and any two algorithms, respectively (Supplementary Figure S1). Second structures of the stem-loop miRNA precursors were predicted by RNAfold (49). Furthermore, the target genes of these putative miRNAs were predicted by two well-known algorithms: TargetScan (50) and miRanDa (51). All information was stored in a local MySQL server.

miRNA quantification, isomiR identification and arm switching discovery

As shown in Figure 1, predicted novel miRNAs were combined with known miRNAs from miRBase R19. All pre-processed datasets were mapped to the combined miRNA list using Bowtie with options: -a -v 1 -S -f –norc, and then alignment results were produced in a BAM file format by SAMtools (52). The BAM files were

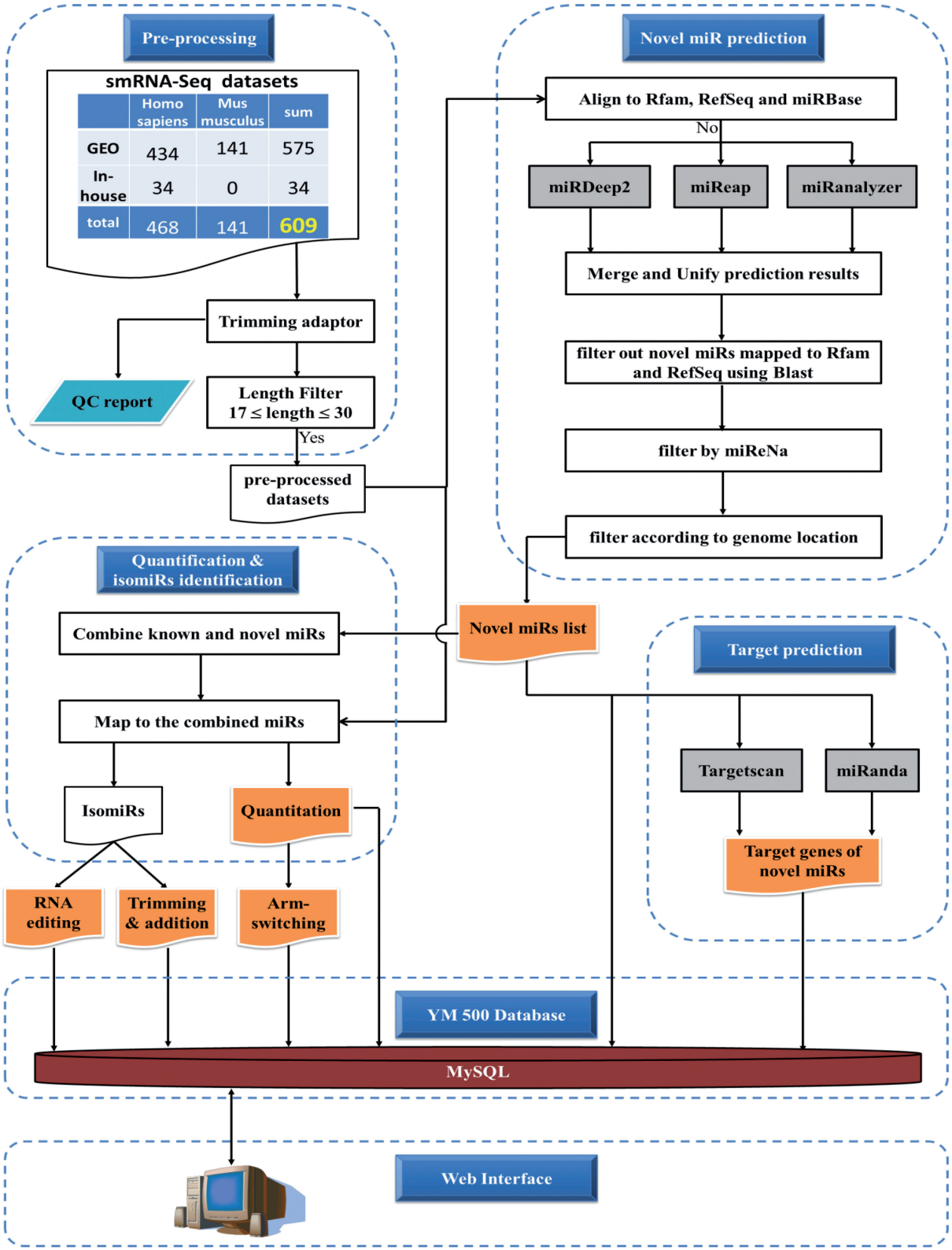

Figure 1. Schematic representation of data processing.

processed by in-house JAVA software for miRNA quantification and isomiR identification. These analysis results were then stored in a local MySQL database. The arm switching events between two groups of samples were determined by two criteria: one was that the averaged RPKM (Reads Per Kilobase of transcript per Million mapped reads) value of a miRNA must be larger than 100. The second was that the ratios of two arms (5p/3p)

in two groups must be significantly different according to the Student T-test with a P-value < 0.05 performed with the R language.

Validation of miRNA expression via stem-loop real-time PCR

Quantification of mature miRNAs was validated by a stem-loop real-time RT-qPCR system performed as previously described (53). Samples ranging from 100 ng to 1 μg of total RNA were used to perform reverse transcription (RT) using the RevertAid™ Reverse transcriptase kit (K1622; Fermentas, Glen Burnie, Maryland, USA) as directed by the manufacturer. Real-time PCR reactions were performed using Maxima™ SYBR Green qPCR Master Mix (K0222; Fermentas, Glen Burnie, Maryland, USA), and the specific products were detected and analysed using the StepOne™ sequence detector (Applied Biosystems, USA). Primers were designed on the basis of the sequenced miRNAs by using FastPCR (54). The miRNA expression data were normalized against U6 small nuclear RNA.

WEB INTERFACE

YM500 provides four interactive query interfaces (Expression, Novel miRNAs, isomiRs and arm-switching) and various graphical visualization pages to present the analysis results of hundreds of smRNA-Seq datasets.

Expression

YM500 allows expression visualization according to a user's customized selections. This feature helps users to select miRNAs according to ID lists or miRNA cluster definitions (Supplementary Figure S2A). For a query regarding a single miRNA, we provide the histogram expression in all samples and the expression profiles by tissue type of the specific miRNA. For a query regarding multiple miRNAs, samples can be selected according to the annotation (Supplementary Figure S2B). A recheck page (Supplementary Figure S2C) helps users to select the specific miRNA and samples for heatmap visualization (Figure 2A). A download link for the normalized expression data of selected samples and miRNAs is also provided for further analysis. Differential expression of 114 miRNAs was confirmed by RT-qPCR. The Pearson's correlation coefficient, R, between NGS analysis and RT-qPCR results was 0.84 (Figure 2B).

Novel miRNAs

Whenever researchers identify potentially novel miRNAs, they can search for existing evidence of the novel miRNAs among the hundreds of samples in YM500. Novel miRNAs can be searched for according to the exact sequence of a mature miRNA or genomic location. Figure 3A illustrates the provided mature and precursor miRNA information, including the prediction algorithms, the predicted target genes, the expression profiles, the RNA secondary structures and the hyperlinks to three commonly used genome browsers. A density plot of reads that mapped to the stem-loop precursor indicates the percentage of reads overlapping the loci (Figure 3B). This provides an overview of length heterogeneity for a miRNA. Figure 3C shows a view of the deep sequencing reads and illustrates all sequences that map to the same novel miRNA. In the same figure, we also provide read numbers and numbers of datasets/samples in which a novel miRNA sequence was found. Reads can be filtered according to the number of mismatches to the hairpin sequence, the read count and the number of datasets/samples. Furthermore, each specific sequence (for example, the sequence in the rectangle in Figure 3C) has a hyperlink to a page (shown in Figure 3D) which contains the detailed information for the sequence, including the expression histogram, the raw counts and the RPM (Reads Per Million mapped reads) in each sample.

IsomiRs

This section helps researchers to find the isomiRs of the known miRNAs in miRBase. As shown in Figure 4A, users can define the criteria of isomiRs according to number of mismatch, number of read counts, number of expressed samples and isomiR types (trimming or addition at 5′/3′ end). Figure 4B illustrates the information of edited sites, which are determined by the editing rate in Figure 4A. Editing information is also compared to dbSNP (Build 135). As shown in Figure 4B, both editing and an SNP were found at the 23rd base of hsa-miR-211-5p. However, the type of nucleotide alteration is different, indicating that such nucleotide substitution is a putative editing event rather than a common variant. All of the isomiRs defined in Figure 4A are detailed in Figure 4C, which indicates the mismatch sites, number of reads and number of samples containing the isomiR. Similar to the view of NGS data in the *Novel miRNAs* part (Figure 3C), each sequence in Figure 4C has a hyperlink to a page, shown in Figure 3D, which summarizes the details of the specific isomiR. Figure 4D shows the summary information of all isomiRs of a specific mature miRNA by a sequence logo format (where the height of each character is proportional to the total read counts of a miRNA). Figure 4E is similar to Figure 4D, but the height of each character is normalized to the read counts in each base and indicates the editing rate in each base. Similar to Figure 3B, Figure 4F is a density plot showing the distribution of reads in a mature miRNA and illustrates the length heterogeneity.

Arm switching

As far as arm switching is concerned, YM500 provides two ways to investigate this phenomenon. YM500 allows users to select a specific precursor miRNA and profiles the expression of two arms between samples and tissues (Figure 5A). This helps researcher to quickly view arm switching events in a specific miRNA species. Another method of illustration allows users to select two groups of samples, according to annotations for the database, and YM500 will identify precursor miRNAs whose dominant expression switches from one arm to the other between the two groups (Figure 5B).

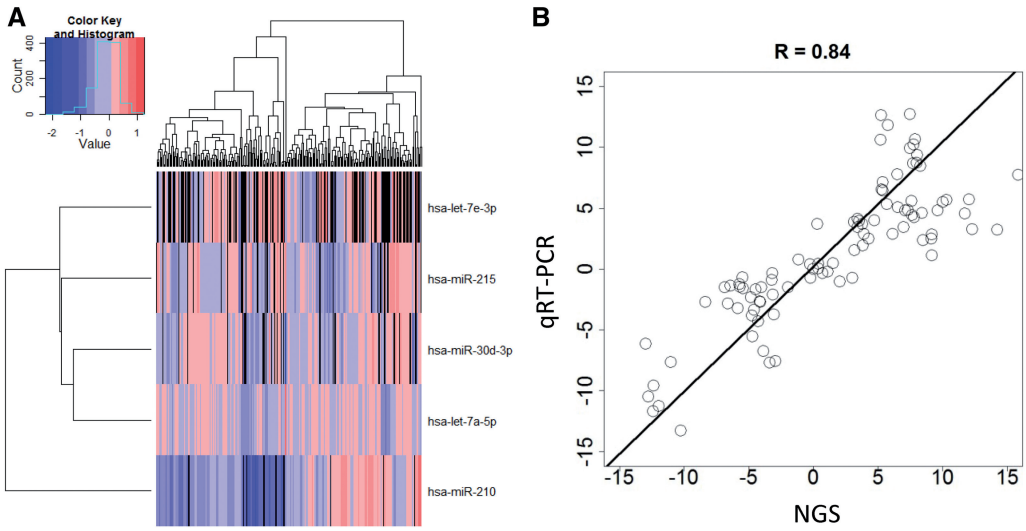

Figure 2. miRNA expression. (**A**) The heatmap visualization for the expression of the selected miRNAs across samples. (**B**) The comparison of NGS and RT-qPCR in quantification of mature miRNAs.

DISCUSSION

YM500 integrates the analysis results of miRNA profiling/quantification, isomiR detection, novel miRNA prediction and arm switching identification. The reliability of *in silico* data were tested via experimental validation on in-house samples or by paper survey. For quantification, our miRNA profiling results were highly consistent to those of RT-qPCR, with a Person's correlation coefficient 0.84 (Figure 2D). For novel miRNA discovery, we used multi-algorithms to explore as many putative miRNAs as possible and adopted several filtration steps to reduce false positive discoveries. The definition of 'novel miRNA' in YM500 is with respect to miRBase and we are collecting the information of 'novel miRNA' claimed by other references as another source of evidences but this task is undergoing. A dozen of miRNAs have been validated in our lab for their existence and expression patterns by RT-qPCR (Supplementary Table S2). Using 'NM_hsa_1300' as an example (Figure 3), wetlab evidence such as the RT-qPCR melting curve proved its existence (Supplementary Figure S3). However, RT-qPCR can prove the 'existence' of putative miRNAs. We use Ago1/2-mediated RNA-immunoprecipitation (RNA-IP) plus further sequencing as another line of evidence, and found that >1000 novel miRNAs in YM500 are indeed associated with the RNA-induced silencing complex (RISC) (unpublished data). For isomiRs, a U-to-G substitution in the ninth base of mmu-let-7a-5p (32) is discovered in YM500. Such substitution events (putative editing event) may result in a significant increase in stability of down-regulated targets (32). The second example is that we also discovered three reported A-to-I putative editing events in three miRNAs, which were the 7th, 8th and 9th bases of the mature product, and edited in 25%, 18% and 11% of the reads in (55), indicating that additional variability is tolerated in the functionally important seed region.

As for an example of arm switching, for hsa-mir-154 (Figure 5A), Cloonan *et al.* have reported that the expression of the two arms would be switched in different tissues (14). They demonstrated the switch between expression dominance from the 3p arm (ovary) to the 5p arm (brain and placenta). It has been shown that alternative mature miRNAs produced from the same precursor have different targeting properties and therefore different biological functions (56). Hence, the changes in arm choice of hsa-mir-154 might have significant functional consequences. The expression of a specific arm may dominate some tissue-specific functions. Besides, hsa-mir-144 has been reported as a cancer/disease marker in several studies (57–59) but our arm switching results show that it might have significant tissue-specificity. It may need to be further investigated the mechanisms of hsa-mir-144 which are related to cancer and tissue-specificity.

The advantage of using smRNA-Seq for miRNA researches is that smRNA-Seq can discover novel miRNAs and isomiRs. At this point, the number and functions of isomiRs remain unclear, but it has been reported that they might be quite prevalent in creatures. YM500 allows researchers to define the criteria of isomiRs, as well as providing existing evidence for various isomiRs from hundreds of smRNA-Seq datasets. A representative example of expression patterns and raw read counts in each smRNA-seq dataset is shown in Figure 4. We defined a candidate isomiR by the criteria shown in Figure 4A (allow one mismatch; the read count and sample count are at least 100 and 5, respectively). At the same time, the editing event is defined by the editing rate at least 1% of the total reads (pass the criteria) mapped to hsa-miR-211-5p. (Please note that all of criteria could also be customer-defined). In this case, 13 sites of the canonical miRNAs have isomiRs with mismatch, but there is only one putative editing event in the 23rd base with 834 supporting reads in more than a dozen of datasets (Figure 4B and C). Besides, there are two distinct isomiRs that have the same editing site (the rectangle in Figure 4C). We also provide the detail information of each isomiR via

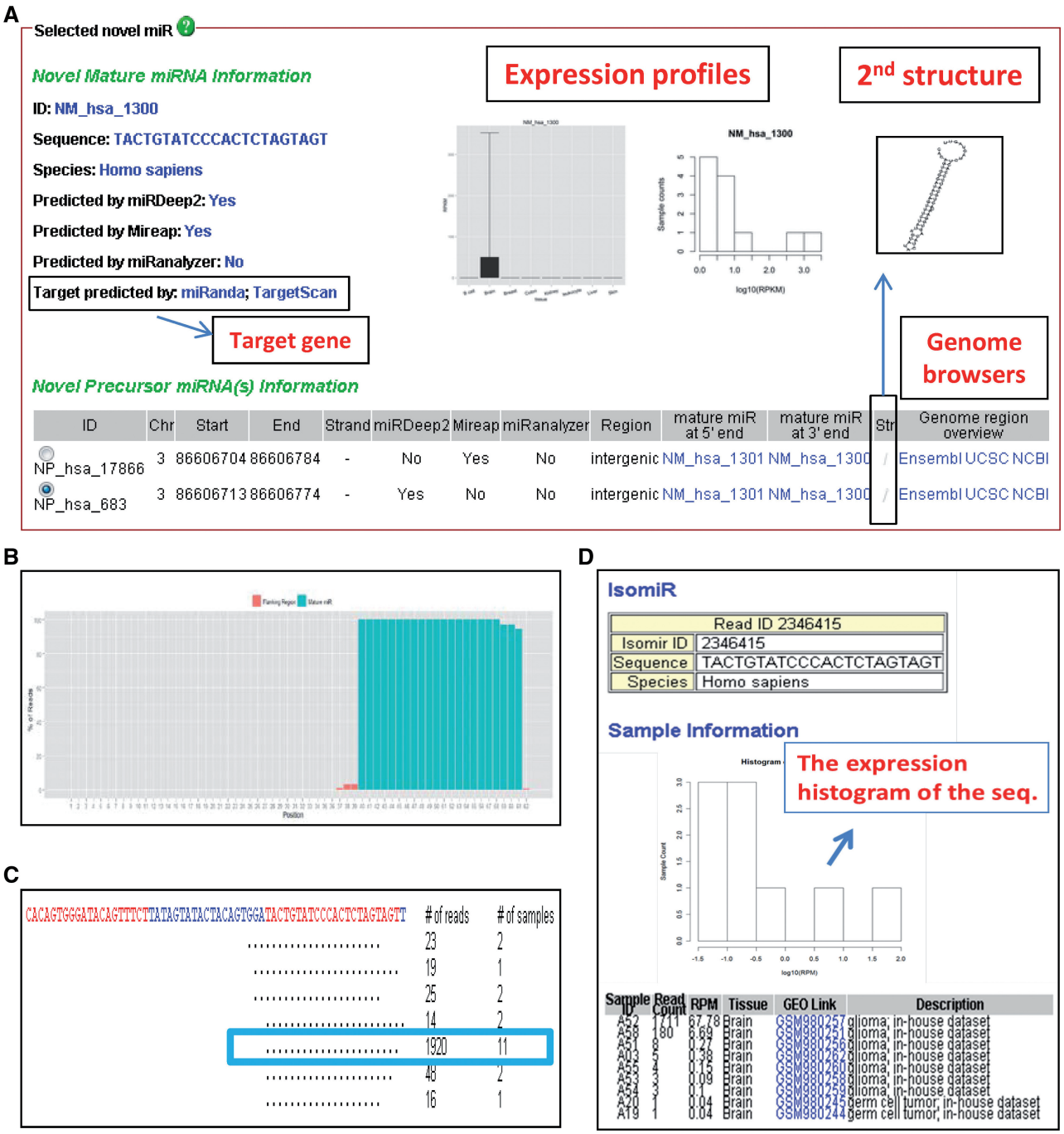

Figure 3. Novel miRNA information. (**A**) The information for a novel mature miRNA (NM_hsa_1300; NM: novel mature miRNAs in YM500) and its precursor miRNA (NP_hsa_17866 and _683; NP: novel precursor miRNAs in YM500). (**B**) A density plot illustrating reads mapping to the putative precursor sequence (NP_hsa_683). Red and green bars indicate the flanking and the mature miRNA regions, respectively. (**C**) The view of deep sequencing reads. Each unique read is mapped to the putative precursor sequence (NP_hsa_683), with the putative mature sequence (NM_hsa_1300) highlighted in yellow. Dots indicate 'perfect match.' Numbers on the right show the read counts of each unique sequence and the number of samples in which this sequence was found. (**D**) The detailed information of the sequence labeled in the rectangle of (C). The expression histogram (middle) and the read counts and RPM of the sequence in the corresponding datasets (bottom) are shown.

hyperlinks on web interface. These isomiRs still need extensive wetlab validations to prove their existence and functionality. Especially, the editing events need genomic evidences from the same sample to rule out novel SNPs or mutations. However, YM500 is a screening tool to help researchers reduce the numbers of candidate isomiRs and could be severed as a line of evidences for their existence. We believe these customer-defined criteria and selections would help researchers to exclude most sequencing errors and other artifacts.

The same holds true for novel miRNAs. When a small RNA is consistently detected in various datasets, it does

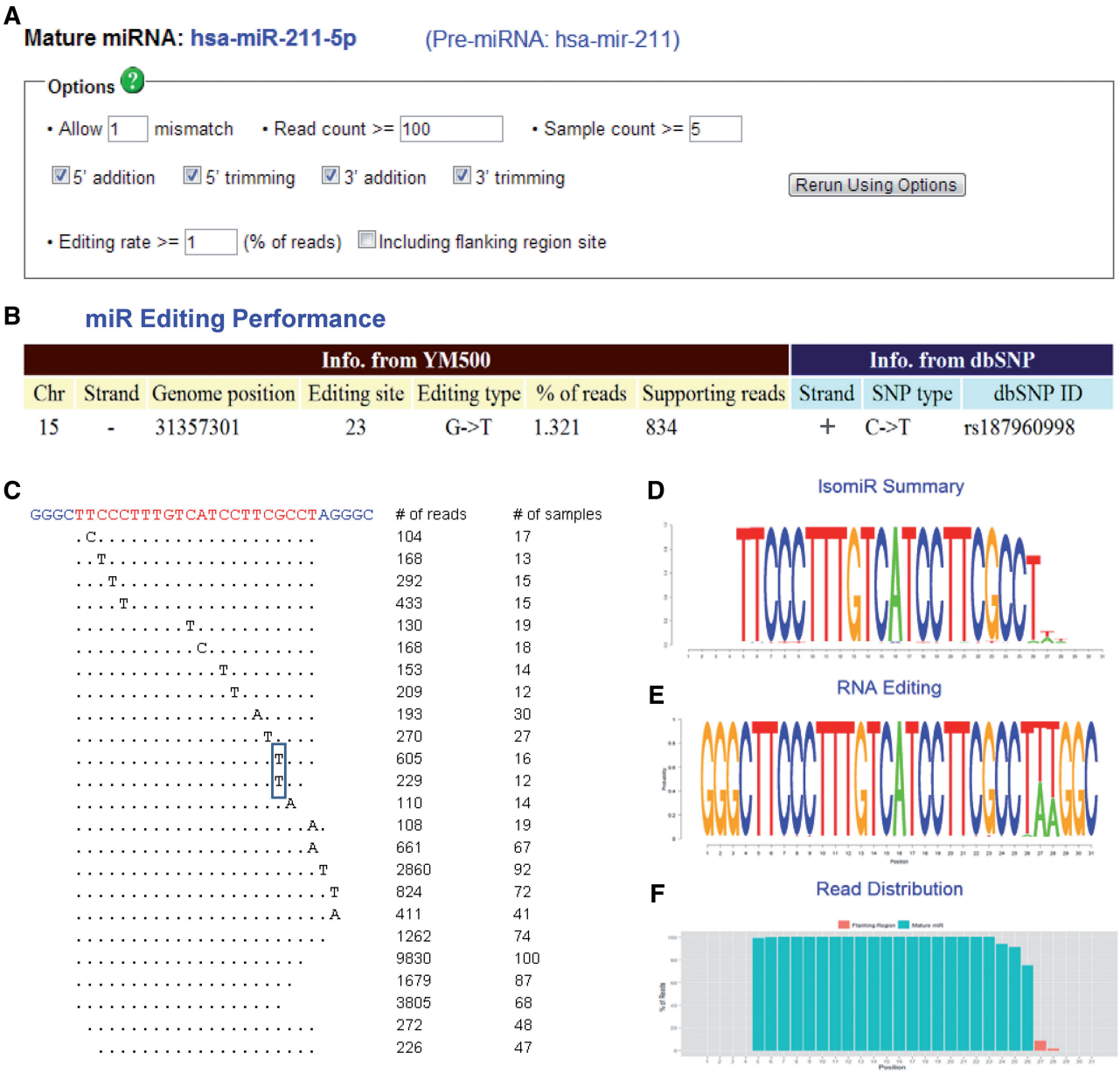

Figure 4. IsomiR summarization. (**A**) A panel to define the criteria of isomiRs of a mature miRNA (e.g. hsa-miR-211-5p) from the precursor (hsa-mir-211). (**B**) The information of edited sites determined by the editing rate defined in (A). Editing information is also compared to that in dbSNP (Build 135). (**C**) The view of deep sequencing reads mapped to the sequence which contains the mature miRNA (hsa-miR-211-5p) and several flanking bases from the precursor (hsa-mir-211). Each unique isomiR is mapped to the sequence (top), with the mature sequence highlighted (yellow). Numbers on the right show the read count of each unique isomiR and the number of samples in which the isomiR was found. The rectangle indicates the putative editing site in (B). (**D–E**) The summary information for all isomiRs of a specific mature miRNA is illustrated in a sequence logo format. The number of the *y*-axis corresponds to the order of the sequencing on the top of (C). The height of each character is proportional to the total read counts (D) or to the read counts in each base (E) of the mature miRNA (hsa-miR-211-5p). (**F**) The density plot of reads mapping to the sequence on the top of (C). The red and green bars indicate the flanking and the mature miRNA regions, respectively.

constitute autonomous evidence that it is prevalent and thus the likely result of a specific biogenesis (6). As shown in Figure 3, for a putative novel miRNA, YM500 provides the expression profiles, the prediction algorithms, the sequences, the counts mapping to the miRNA, the secondary structure, etc. This information helps researchers to evaluate the prediction result. For example, according to the suggestions of Kozomara and Griffiths-Jones (45), the pattern of reads focusing on the mature region (Figure 3B) supports a high-confidence miRNA annotation with multiple reads (10–20 as cutoffs) from independent datasets. In contrast, the pattern of reads overlapping the sequence of a putative miRNA does not support the annotation of a miRNA, with multiple offset reads distributed across the locus (45). Besides, most reads mapping to a given mature miRNA annotation should have the same 5′ end whereas the 3′ end may be significantly more variable

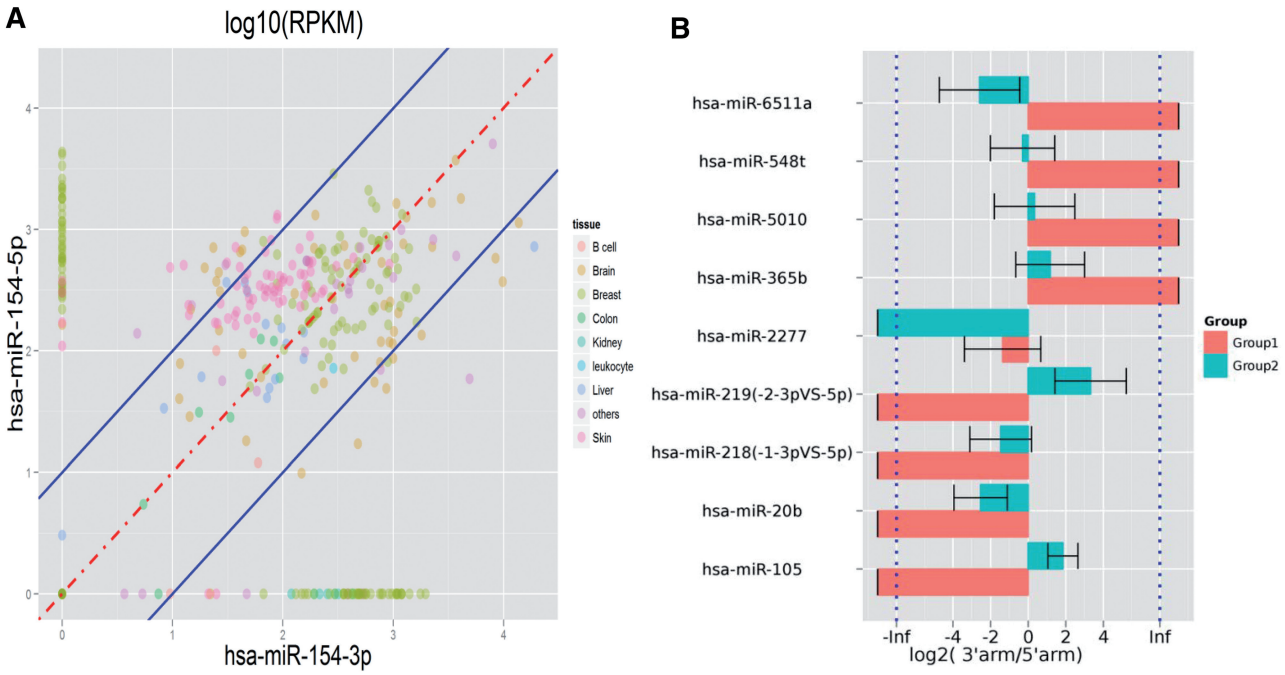

Figure 5. Arm switching. (**A**) The expression profiles of two arms of hsa-mir-154 between samples and tissues. (**B**) The precursor miRNAs with arm switching event, identified by YM500, between two groups of customer-defined samples.

(45). YM500 helps researchers to check these characteristics via the presentation page (Figure 3). For isomiRs and novel miRNAs, YM500 provides existing evidence from hundreds of smRNA-Seq datasets. Whenever researchers find interesting results (such as some specific isomiRs and novel miRNAs) in their own datasets, they can validate their results in YM500.

In comparison to other databases related to smRNA-seq, including miRBase, deepBase (60) and the isomiRs database of Lee *et al.* (24), YM500 analyse miRNA-related information from many dimensions, and datasets included in meta-analysis are more abundant. The deepBase database contains 185 smRNA-Seq datasets for seven organisms and is also a platform for annotating and discovering small and long non-coding RNAs (miRNAs, siRNAs, piRNAs, etc.) from next-generation sequencing data. The isomiRs database of Lee *et al.* contains only 18 smRNA-Seq datasets for two organisms and lists only the isomiR information in a single sample per query. Although miRBase covers much more species, miRBase does not include expression profiles of known miRNAs, RNA editing, arm switching or miRNA prediction results. For novel miRNA, YM500 adopts four different algorithms and provides target prediction, expression profiles of novel miRNA and hyperlinks to genome browsers. deepBase lists only the results of novel miRNA predicted by miRDeep, and no other additional information is provided. Besides, there are only 79, 9 and 5 human smRNA-seq datasets in miRBase, deepBase and the isomiRs database of Lee *et al.*, respectively. YM500 contains 468 human datasets (from public databases or in-house), which is the most comprehensive collection so far. Another unique part of YM500 is that our interactive web interface and customer-defined criteria also help researchers to retrieve these four types of analysis results. Finally, to our best knowledge, there is no other resource providing arm switching information. Comparing with these databases, YM500 provides a flexible web interface, more enhanced resolution and novel findings owing to the integrated pipelines for miRNA (including miRNA expression, IsomiRs, novel miRNAs and arm-switching) and the large number of smRNA-Seq datasets of various tissue/cell types.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2 and Supplementary Figures 1–3.

FUNDING

Funding for open access charge: National Science Council [NSC98-2320-B-010-020-MY3, NSC99-2320-B-350-003-MY3, NSC100-2627-B-010-007, NSC100-2314-B-075-066-MY2 and NSC101-2320-B-010-059-MY3]; Taipei Veterans General Hospital [V101E2-008 and Cancer Excellence Center Plan, DOH101-TD-C-111-007]; National Research Program for Biopharmaceuticals [DOH101-TD-PB-111-TM007]; National Yang-Ming University via the Ministry of Education, Aim for the Top University Plan; UST-UCSD International Center for Excellence in Advanced Bioengineering sponsored by the Taiwan NSC I-RiCE Program [NSC100-2911-I-009-101, in part].

Conflict of interest statement. None declared.

REFERENCES

1. Zhou,L., Li,X., Liu,Q., Zhao,F. and Wu,J. (2011) Small RNA transcriptome investigation based on next-generation sequencing technology. *J. Genet. Genomics*, **38**, 505–513.
2. Morin,R.D., O'Connor,M.D., Griffith,M., Kuchenbauer,F., Delaney,A., Prabhu,A.L., Zhao,Y., McDonald,H., Zeng,T., Hirst,M. *et al.* (2008) Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.*, **18**, 610–621.
3. Barrett,T., Troup,D.B., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M. *et al.* (2011) NCBI GEO: archive for functional genomics data sets–10 years on. *Nucleic Acids Res.*, **39**, D1005–D1010.
4. Parkinson,H., Sarkans,U., Kolesnikov,N., Abeygunawardena,N., Burdett,T., Dylag,M., Emam,I., Farne,A., Hastings,E., Holloway,E. *et al.* (2011) ArrayExpress update–an archive of microarray and high-throughput sequencing-based functional genomics experiments. *Nucleic Acids Res.*, **39**, D1002–D1004.
5. Li,Y., Zhang,Z., Liu,F., Vongsangnak,W., Jing,Q. and Shen,B. (2012) Performance comparison and evaluation of software tools for microRNA deep-sequencing data analysis. *Nucleic Acids Res.*, **40**, 4298–4305.
6. Friedlander,M.R., Mackowiak,S.D., Li,N., Chen,W. and Rajewsky,N. (2012) miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.*, **40**, 37–52.
7. Hackenberg,M., Rodriguez-Ezpeleta,N. and Aransay,A.M. (2011) miRanalyzer: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments. *Nucleic Acids Res.*, **39**, W132–W138.
8. Hendrix,D., Levine,M. and Shi,W. (2010) miRTRAP, a computational method for the systematic identification of miRNAs from high throughput sequencing data. *Genome Biol.*, **11**, R39.
9. Mathelier,A. and Carbone,A. (2010) MIReNA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data. *Bioinformatics*, **26**, 2226–2234.
10. Zhu,E., Zhao,F., Xu,G., Hou,H., Zhou,L., Li,X., Sun,Z. and Wu,J. (2010) mirTools: microRNA profiling and discovery based on high-throughput sequencing. *Nucleic Acids Res.*, **38**, W392–W397.
11. Huang,P.J., Liu,Y.C., Lee,C.C., Lin,W.C., Gan,R.R., Lyu,P.C. and Tang,P. (2010) DSAP: deep-sequencing small RNA analysis pipeline. *Nucleic Acids Res.*, **38**, W385–W391.
12. Ronen,R., Gan,I., Modai,S., Sukacheov,A., Dror,G., Halperin,E. and Shomron,N. (2010) miRNAkey: a software for microRNA deep sequencing analysis. *Bioinformatics*, **26**, 2615–2616.
13. Chen,X., Li,Q., Wang,J., Guo,X., Jiang,X., Ren,Z., Weng,C., Sun,G., Wang,X., Liu,Y. *et al.* (2009) Identification and characterization of novel amphioxus microRNAs by Solexa sequencing. *Genome Biol.*, **10**, R78.
14. Cloonan,N., Wani,S., Xu,Q., Gu,J., Lea,K., Heater,S., Barbacioru,C., Steptoe,A.L., Martin,H.C., Nourbakhsh,E. *et al.* (2011) MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol.*, **12**, R126.
15. Danecek,P., Nellaker,C., McIntyre,R.E., Buendia-Buendia,J.E., Bumpstead,S., Ponting,C.P., Flint,J., Durbin,R., Keane,T.M. and Adams,D.J. (2012) High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. *Genome Biol.*, **13**, r26.
16. Ekdahl,Y., Farahani,H.S., Behm,M., Lagergren,J. and Ohman,M. (2012) A-to-I editing of microRNAs in the mammalian brain increases during development. *Genome Res.*, **22**, 1477–1487.
17. Vesely,C., Tauber,S., Sedlazeck,F.J., von Haeseler,A. and Jantsch,M.F. (2012) Adenosine deaminases that act on RNA induce reproducible changes in abundance and sequence of embryonic miRNAs. *Genome Res.*, **22**, 1468–1476.
18. Wyman,S.K., Knouf,E.C., Parkin,R.K., Fritz,B.R., Lin,D.W., Dennis,L.M., Krouse,M.A., Webster,P.J. and Tewari,M. (2011) Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. *Genome Res.*, **21**, 1450–1461.
19. Chen,B., Zhang,B., Luo,H., Yuan,J., Skogerbo,G. and Chen,R. (2012) Distinct microRNA subcellular size and expression patterns in human cancer cells. *Int. J. Cell Biol.*, **2012**, 672462.
20. Guduric-Fuchs,J., O'Connor,A., Cullen,A., Harwood,L., Medina,R.J., O'Neill,C.L., Stitt,A.W., Curtis,T.M. and Simpson,D.A. (2012) Deep sequencing reveals predominant expression of miR-21 amongst the small non-coding RNAs in retinal microvascular endothelial cells. *J. Cell. Biochem.*, **113**, 2098–2111.
21. Guo,L., Li,H., Liang,T., Lu,J., Yang,Q., Ge,Q. and Lu,Z. (2012) Consistent isomiR expression patterns and 3′ addition events in miRNA gene clusters and families implicate functional and evolutionary relationships. *Mol. Biol. Rep.*, **39**, 6699–6706.
22. Burroughs,A.M., Kawano,M., Ando,Y., Daub,C.O. and Hayashizaki,Y. (2012) pre-miRNA profiles obtained through application of locked nucleic acids and deep sequencing reveals complex 5′/3′ arm variation including concomitant cleavage and polyuridylation patterns. *Nucleic Acids Res.*, **40**, 1424–1437.
23. Zhou,H., Arcila,M.L., Li,Z., Lee,E.J., Henzler,C., Liu,J., Rana,T.M. and Kosik,K.S. (2012) Deep annotation of mouse iso-miR and iso-moR variation. *Nucleic Acids Res.*, **40**, 5864–5875.
24. Lee,L.W., Zhang,S., Etheridge,A., Ma,L., Martin,D., Galas,D. and Wang,K. (2010) Complexity of the microRNA repertoire revealed by next-generation sequencing. *RNA*, **16**, 2170–2180.
25. Newman,M.A., Mani,V. and Hammond,S.M. (2011) Deep sequencing of microRNA precursors reveals extensive 3′ end modification. *RNA*, **17**, 1795–1803.
26. Voellenkle,C., van Rooij,J., Guffanti,A., Brini,E., Fasanaro,P., Isaia,E., Croft,L., David,M., Capogrossi,M.C., Moles,A. *et al.* (2012) Deep-sequencing of endothelial cells exposed to hypoxia reveals the complexity of known and novel microRNAs. *RNA*, **18**, 472–484.
27. Humphreys,D.T., Hynes,C.J., Patel,H.R., Wei,G.H., Cannon,L., Fatkin,D., Suter,C.M., Clancy,J.L. and Preiss,T. (2012) Complexity of murine cardiomyocyte miRNA biogenesis, sequence variant expression and function. *PLoS One*, **7**, e30933.
28. Ebhardt,H.A., Tsang,H.H., Dai,D.C., Liu,Y., Bostan,B. and Fahlman,R.P. (2009) Meta-analysis of small RNA-sequencing errors reveals ubiquitous post-transcriptional RNA modifications. *Nucleic Acids Res.*, **37**, 2461–2470.
29. Pantano,L., Estivill,X. and Marti,E. (2010) SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells. *Nucleic Acids Res.*, **38**, e34.
30. Blow,M.J., Grocock,R.J., van Dongen,S., Enright,A.J., Dicks,E., Futreal,P.A., Wooster,R. and Stratton,M.R. (2006) RNA editing of human microRNAs. *Genome Biol.*, **7**, R27.
31. Kawahara,Y., Zinshteyn,B., Sethupathy,P., Iizasa,H., Hatzigeorgiou,A.G. and Nishikura,K. (2007) Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science*, **315**, 1137–1140.
32. Reid,J.G., Nagaraja,A.K., Lynn,F.C., Drabek,R.B., Muzny,D.M., Shaw,C.A., Weiss,M.K., Naghavi,A.O., Khan,M., Zhu,H. *et al.* (2008) Mouse let-7 miRNA populations exhibit RNA editing that is constrained in the 5′-seed/cleavage/anchor regions and stabilize predicted mmu-let-7a:mRNA duplexes. *Genome Res.*, **18**, 1571–1581.
33. Alon,S., Mor,E., Vigneault,F., Church,G.M., Locatelli,F., Galeano,F., Gallo,A., Shomron,N. and Eisenberg,E. (2012) Systematic identification of edited microRNAs in the human brain. *Genome Res.*, **22**, 1533–1540.
34. Han,B.W., Hung,J.H., Weng,Z., Zamore,P.D. and Ameres,S.L. (2011) The 3′-to-5′ exoribonuclease Nibbler shapes the 3′ ends of microRNAs bound to Drosophila Argonaute1. *Curr. Biol.*, **21**, 1878–1887.
35. Liu,N., Abe,M., Sabin,L.R., Hendriks,G.J., Naqvi,A.S., Yu,Z., Cherry,S. and Bonini,N.M. (2011) The exoribonuclease Nibbler controls 3′ end processing of microRNAs in Drosophila. *Curr. Biol.*, **21**, 1888–1893.
36. Katoh,T., Sakaguchi,Y., Miyauchi,K., Suzuki,T., Kashiwabara,S. and Baba,T. (2009) Selective stabilization of mammalian microRNAs by 3′ adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. *Genes Dev.*, **23**, 433–438.

37. Jones,M.R., Quinton,L.J., Blahna,M.T., Neilson,J.R., Fu,S., Ivanov,A.R., Wolf,D.A. and Mizgerd,J.P. (2009) Zcchc11-dependent uridylation of microRNA directs cytokine expression. *Nat. Cell Biol.*, **11**, 1157–1163.
38. Burroughs,A.M., Ando,Y., de Hoon,M.J., Tomaru,Y., Nishibu,T., Ukekawa,R., Funakoshi,T., Kurokawa,T., Suzuki,H., Hayashizaki,Y. *et al.* (2010) A comprehensive survey of 3′ animal miRNA modification events and a possible role for 3′ adenylation in modulating miRNA targeting effectiveness. *Genome Res.*, **20**, 1398–1410.
39. Griffiths-Jones,S., Hui,J.H., Marco,A. and Ronshaugen,M. (2011) MicroRNA evolution by arm switching. *EMBO Rep.*, **12**, 172–177.
40. Marco,A., Hui,J.H., Ronshaugen,M. and Griffiths-Jones,S. (2010) Functional shifts in insect microRNA evolution. *Genome Biol. Evol.*, **2**, 686–696.
41. Li,S.C., Liao,Y.L., Ho,M.R., Tsai,K.W., Lai,C.H. and Lin,W.C. (2012) miRNA arm selection and isomiR distribution in gastric cancer. *BMC Genomics*, **13(Suppl. 1)**, S13.
42. Li,S.C., Liao,Y.L., Chan,W.C., Ho,M.R., Tsai,K.W., Hu,L.Y., Lai,C.H., Hsu,C.N. and Lin,W.C. (2011) Interrogation of rabbit miRNAs and their isomiRs. *Genomics*, **98**, 453–459.
43. Grimson,A., Srivastava,M., Fahey,B., Woodcroft,B.J., Chiang,H.R., King,N., Degnan,B.M., Rokhsar,D.S. and Bartel,D.P. (2008) Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature*, **455**, 1193–1197.
44. Wang,W.C., Lin,F.M., Chang,W.C., Lin,K.Y., Huang,H.D. and Lin,N.S. (2009) miRExpress: analyzing high-throughput sequencing data for profiling microRNA expression. *BMC Bioinform.*, **10**, 328.
45. Kozomara,A. and Griffiths-Jones,S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, **39**, D152–D157.
46. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
47. Gardner,P.P., Daub,J., Tate,J., Moore,B.L., Osuch,I.H., Griffiths-Jones,S., Finn,R.D., Nawrocki,E.P., Kolbe,D.L., Eddy,S.R. *et al.* (2011) Rfam: Wikipedia, clans and the "decimal" release. *Nucleic Acids Res.*, **39**, D141–D145.
48. Pruitt,K.D., Tatusova,T., Brown,G.R. and Maglott,D.R. (2012) NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.*, **40**, D130–D135.
49. Zuker,M. and Stiegler,P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.*, **9**, 133–148.
50. Lewis,B.P., Shih,I.H., Jones-Rhoades,M.W., Bartel,D.P. and Burge,C.B. (2003) Prediction of mammalian microRNA targets. *Cell*, **115**, 787–798.
51. Enright,A.J., John,B., Gaul,U., Tuschl,T., Sander,C. and Marks,D.S. (2003) MicroRNA targets in Drosophila. *Genome Biol.*, **5**, R1.
52. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
53. Chen,C., Ridzon,D.A., Broomer,A.J., Zhou,Z., Lee,D.H., Nguyen,J.T., Barbisin,M., Xu,N.L., Mahuvakar,V.R., Andersen,M.R. *et al.* (2005) Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.*, **33**, e179.
54. Kalendar,R., Lee,D. and Schulman,A.H. (2011) Java web tools for PCR, in silico PCR, and oligonucleotide assembly and analysis. *Genomics*, **98**, 137–144.
55. Parts,L., Hedman,A.K., Keildson,S., Knights,A.J., Abreu-Goodger,C., van de Bunt,M., Guerra-Assuncao,J.A., Bartonicek,N., van Dongen,S., Magi,R. *et al.* (2012) Extent, causes, and consequences of small RNA expression variation in human adipose tissue. *PLoS Genet.*, **8**, e1002704.
56. Marco,A., Macpherson,J.I., Ronshaugen,M. and Griffiths-Jones,S. (2012) MicroRNAs from the same precursor have different targeting properties. *Silence*, **3**, 8.
57. Akiyoshi,S., Fukagawa,T., Ueo,H., Ishibashi,M., Takahashi,Y., Fabbri,M., Sasako,M., Maehara,Y., Mimori,K. and Mori,M. (2012) Clinical significance of miR-144-ZFX axis in disseminated tumour cells in bone marrow in gastric cancer cases. *Br. J. Cancer*, **107**, 1345–1353.
58. Gaedcke,J., Grade,M., Camps,J., Sokilde,R., Kaczkowski,B., Schetter,A.J., Difilippantonio,M.J., Harris,C.C., Ghadimi,B.M., Moller,S. *et al.* (2012) The rectal cancer microRNAome - microRNA expression in rectal cancer and matched normal mucosa. *Clin. Cancer Res.*, **18**, 4919–4930.
59. Gu,H., Li,H., Zhang,L., Luan,H., Huang,T., Wang,L., Fan,Y., Zhang,Y., Liu,X., Wang,W. *et al.* (2012) Diagnostic role of microRNA expression profile in the serum of pregnant women with fetuses with neural tube defects. *J. Neurochem.*, **122**, 641–649.
60. Yang,J.H., Shao,P., Zhou,H., Chen,Y.Q. and Qu,L.H. (2010) deepBase: a database for deeply annotating and mining deep sequencing data. *Nucleic Acids Res.*, **38**, D123–D130.